

# APLIKASI PENENTUAN JENIS *PART OF SPEECH* MENGGUNAKAN METODE *N-GRAM* DAN *STRING MATCHING*

<sup>1</sup>Nurzaenab dan <sup>2</sup>Purnawansyah

<sup>1</sup>nurzaenab16@yahoo.co.id, <sup>2</sup>purnawansyah@gmail.com

<sup>12</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muslim Indonesia

## Abstrak

Bahasa Inggris merupakan bahasa ibu dalam skala internasional sebagai alat komunikasi antar negara. Bahasa Inggris memiliki aturan baik dalam hal pengucapan dan penulisan disebut *Grammar* yang membentuk pola-pola. Pola-pola tersebut tersusun oleh setiap kata yang memiliki bentuk-bentuk tersendiri yang disebut *Parts Of Speech*. Bentuk dalam *Parts Of Speech* terbagi dalam delapan bentuk yaitu *Noun* (kata benda), *Pronoun* (kata ganti), *Verb* (kata kerja), *Adjective* (kata sifat), *Adverb* (kata keterangan), *Preposition* (kata depan), *Conjunction* (kata penghubung), *Interjection* (kata seru). Tingkat ingatan manusia tentu berbeda-beda. Ingatan untuk membedakan kata-kata dan pembentukan pola kalimat dalam *part of speech*. Setiap kata akan ditentukan jenis *part of speech*-nya, tergantung dari inputan *user*. Sedangkan pola kalimat akan di tentukan sesuai inputan *user* berdasarkan *part of speech*-nya. Perancangan dilakukan menggunakan metode uni-gram dan String Matching (Knuth Morris Pratt).

**Kata kunci :** *Grammar, N-Gram, Part Of Speech, String Matching*

**Copyright © 2016 -- Jurnal Ilmiah ILKOM -- All rights reserved.**

## 1. Pendahuluan

Bahasa Inggris merupakan bahasa ibu dalam skala internasional sebagai alat komunikasi antar negara. Bahasa Inggris memiliki aturan baik dalam hal pengucapan dan penulisan disebut *Grammar* yang membentuk pola-pola. Pola-pola tersebut tersusun oleh setiap kata yang memiliki bentuk-bentuk tersendiri yang disebut *Parts Of Speech*.

Bentuk dalam *Parts Of Speech* terbagi dalam delapan bentuk yaitu *Noun* (kata benda), *Pronoun* (kata ganti), *Verb* (kata kerja), *Adjective* (kata sifat), *Adverb* (kata keterangan), *Preposition* (kata depan), *Conjunction* (kata penghubung), *Interjection* (kata seru).

Tingkat ingatan manusia tentu berbeda-beda. Ingatan untuk membedakan kata-kata dan pembentukan pola kalimat dalam *part of speech*. Setiap kata akan ditentukan jenis *part of speech*-nya, tergantung dari inputan *user*. Sedangkan pola kalimat akan di tentukan sesuai inputan *user* berdasarkan *part of speech*-nya. Hal itu melatarbelakangi penulis untuk merancang sebuah aplikasi untuk menentukan kata-kata dan pola kalimat dalam bentuk *Parts Of Speech*. Metode yang digunakan adalah metode uni-gram dan *String Matching* (Knuth Morris Pratt).

Tujuan penelitian adalah menghasilkan aplikasi yang dapat mengklasifikasikan kata-kata dan pola kalimat kedalam bentuk *Parts Of Speech* menggunakan metode uni-gram dan String Matching (Knuth Morris Pratt). Penelitian ini memiliki beberapa batasan permasalahan yaitu: hanya menentukan kata-kata dan pola kalimat dalam bentuk *Parts Of Speech*.

## 2. Landasan Teori

### 2.1 Part Of Speech

*Part Of Speech* adalah sebuah pengklasifikasian kata sesuai dengan bentuk kata. Bentuk dalam *Parts Of Speech* terbagi dalam delapan bentuk yaitu *Noun* (kata benda), *Pronoun* (kata ganti), *Verb* (kata kerja), *Adjective* (kata sifat), *Adverb* (kata keterangan), *Preposition* (kata depan), *Conjunction* (kata penghubung), *Interjection* (kata seru).

### 2.2 Metode String Matching Knuth Morris Pratt

Algoritma *Knuth Morris Pratt* dikembangkan oleh D. E. Knuth, bersama-sama dengan J. H. Morris dan V. R. Pratt [1][2]. Dengan algoritma KMP waktu pencarian dalam pencocokan *pattern* dan teks dapat berkurang dikarenakan algoritma ini melakukan sejumlah pergeseran lebih jauh sesuai dengan informasi ketidakcocokan *string* antara teks dan *pattern*.

Algoritma *Knuth Morris Pratt* melakukan pencarian string dari arah kiri ke kanan pada teks. Dengan membandingkan antara teks yang ada dengan teks yang ada pada basisdata.

### 2.3 Metode Uni-Gram

Metode Uni-gram adalah salah satu metode n-Gram yang digunakan untuk memprediksi suatu teks. Karakter yang diambil senilai satu karakter kemudian digunakan untuk dicocokkan dalam teks yang terdapat dalam database kata. Contoh: *Computer*. Pengambilan satu karakternya yaitu "C", "o", "m", "p", "u", "t", "e", "r".

Persamaan (1) menunjukkan proses penghitungan skor tiap passage yang dilakukan dengan menjumlahkan semua kemungkinan x-gram yang cocok, dimana  $x \leq$  jumlah kata pada kueri pertanyaan N-gram.

$$\text{Sum} = \sum_{x=1}^n \sum_{y=1}^n (m * 1) \quad (1)$$

Dengan  $m=1$  apabila cocok, dan  $m=0$  untuk selainnya. Untuk skor maksimum dari rumus uni-gram sum pada rumus (3),  $m = 1$  untuk semua nilai dari  $x$  dan  $y$ . Jadi nilai maksimum dari semua kemungkinan uni-gram adalah  $(n*(n+1)/2)$ . Jadi skor yang diperoleh oleh tiap *passage* ditunjukkan oleh persamaan (2).

$$\text{uni-Gram Score} = \frac{n\text{-gram sum}}{(n * \frac{n+1}{2})} \quad (2)$$

### 2.4 UML (Unified Modeling Language)

*Unified Modelling Language* (UML) adalah sebuah bahasa pemodelan sistem dengan visualisasi, merancang dan mendokumentasikan sistem perangkat lunak.

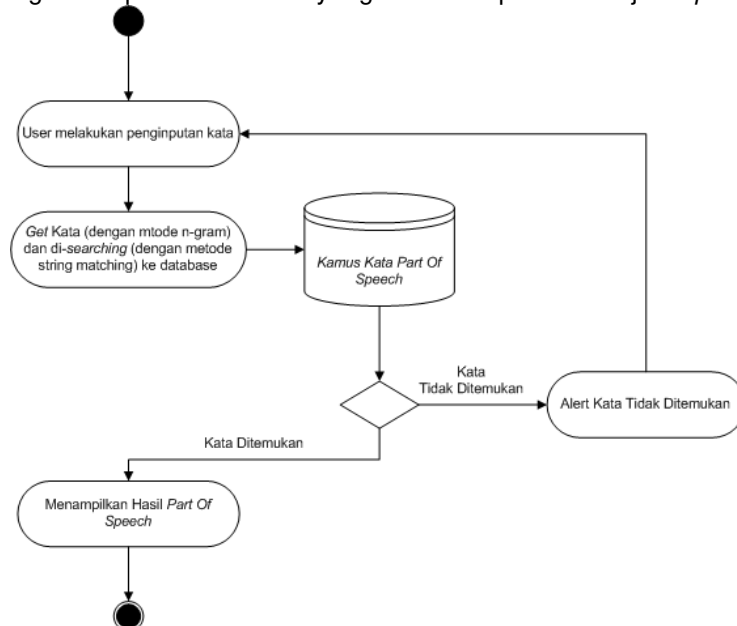
## 3. Metode

### 3.1 Algoritma Metode Uni-Gram dan String Matching Knuth Morris Pratt

Aplikasi Penentuan Jenis *Part Of Speech* adalah aplikasi untuk menentukan kata dan kalimat yang termasuk dalam *part of speech*. Di bawah ini akan diuraikan algoritma implementasi metode n-gram dan string matching (knuth morris pratt).

#### 3.1.1 Algoritma Penentuan Kata

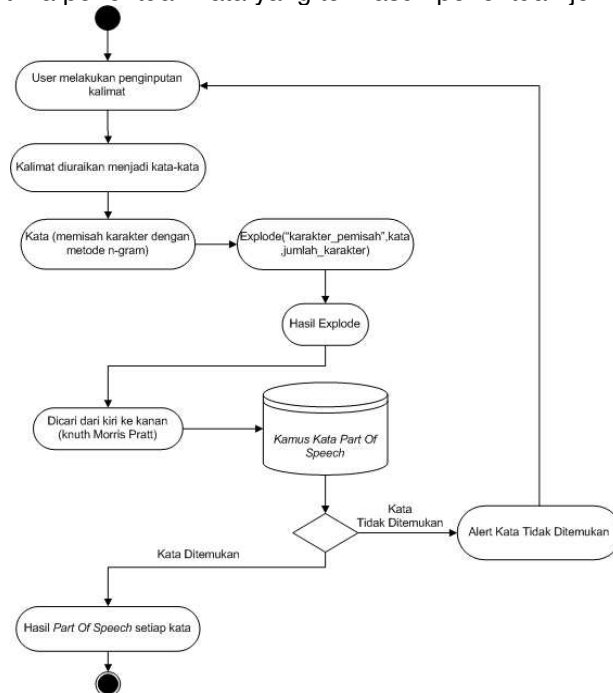
Berikut adalah algoritma penentuan kata yang termasuk penentuan jenis *part of speech*.



Gambar 1 Algoritma penentuan kata yang termasuk *part of speech*.

### 3.1.2 Algoritma Penentuan Kalimat

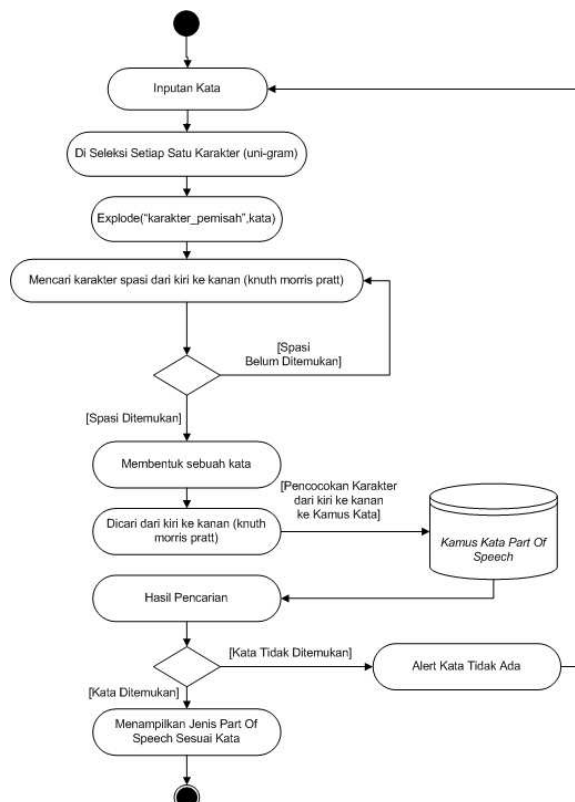
Berikut adalah algoritma penentuan kata yang termasuk penentuan jenis *part of speech*.



Gambar 2 Algoritma penentuan kalimat yang termasuk *part of speech*.

### 3.1.3 Algoritma Penggunaan Metode Uni-Gram dan String Matching

Metode yang digunakan untuk membangun aplikasi penentuan jenis *part of speech* adalah metode uni-gram dan knuth morris pratt. Pengambilan karakter yaitu menggunakan metode uni-gram sedangkan untuk pencarian karakter dan pencocokan karakter pada kamus kata menggunakan metode knuth morris pratt.



Gambar 3 Algoritma Penggunaan Metode Uni-gram dan Knut Morris Pratt.

#### 4. Hasil

Pada implementasi antar muka ini, menjelaskan tentang halaman utama pada aplikasi yang merupakan penghubung dengan sub-sub menu lainnya yang ada pada aplikasi. Berikut adalah penjelasan dari implementasi sistem aplikasi penentuan jenis *part of speech* berbasis web yang dijelaskan di bawah ini.

##### 4.1 Tampilan Menu Informasi

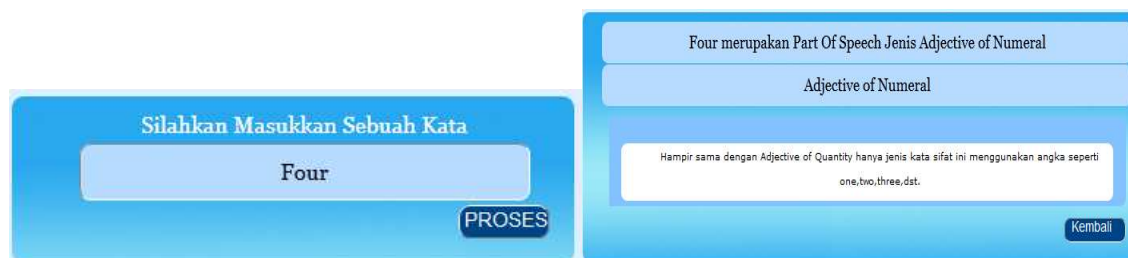
Tampilan menu informasi menampilkan informasi untuk member yang belum mengetahui mengenai aplikasi.



Gambar 4 Tampilan Menu Informasi

##### 4.2 Tampilan Penentuan kata

Tampilan penentuan kata menjelaskan aktifitas member dalam menggunakan aplikasi. Member melakukan penginputan kata kemudian diproses oleh *system* dan menampilkan jenis *part of speech*-nya.



Gambar 5 Tampilan Penentuan Kata

##### 4.3 Tampilan Penentuan Kalimat

Tampilan penentuan kalimat menjelaskan aktifitas member dalam menggunakan aplikasi. Member melakukan penginputan kalimat kemudian diproses oleh *system* dan menampilkan bentuk kalimat *part of speech*-nya



Gambar 6 Tampilan Penentuan Kalimat

## 5. Kesimpulan

Berdasarkan hasil analisis perancangan dan pembangunan sistem, maka dapat diambil beberapa kesimpulan sebagai berikut:

1. Aplikasi Penentuan Jenis Part Of Speech adalah aplikasi untuk menentukan sebuah kata atau kata dalam kalimat yang termasuk dalam bentuk part of speech (*Noun, Pronoun, Verb, Adverb, Preposition, Conjuntion, Adjective, Interjection*).
2. Jalannya aplikasi dibangun dengan menggunakan metode N-gram (Uni-gram) dan *String Matching* (Knuth Morris pratt).

## Daftar Pustaka

- [1]. Tio Wibowo, Ardianto Wibowo, Rika Perdana Sari. 2012. *Pembuatan Aplikasi Untuk Mendeteksi Kebenaran Perintah SQL Query Menggunakan Metode Knuth-Morris Pratt (KMP)*. Jurnal of Computer Science.
- [2]. Wulan, Sri. 2011. *Analisis Penerapan String Matching Dalam Komparasi Data Kepesertaan Jaminan Kesehatan Masyarakat (JAMKESMAS)*. Jakarta.